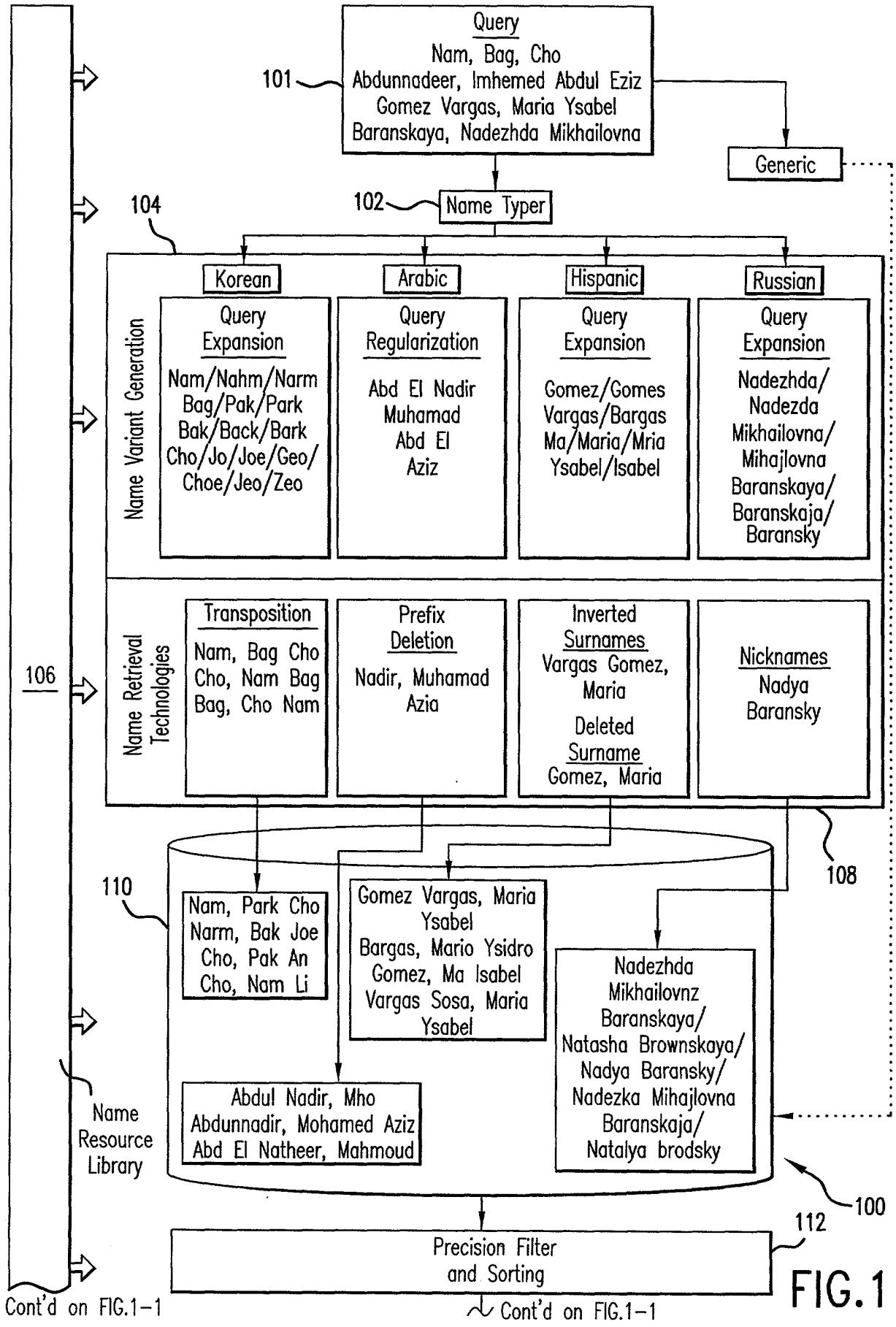
Query
Nam, Bag, Cho
Abdunnadeer, Imhemed Abdul Eziz
Gomez Vargas, Maria Ysabel
Baranskaya, Nadezhda Mikhailovna
101
Generic
102
Name Typer
104
Name Variant Generation
Korean
Arabic
Hispanic
Russian
Query Expansion
Nam/Nahm/Narm
Bag/Pak/Park
Bak/Back/Bark
Cho/Jo/Joe/Geo/
Choe/Jeo/Zeo
Query Regularization
Abd El Nadir
Muhamad
Abd El
Aziz
Query Expansion
Gomez/Gomes
Vargas/Bargas
Ma/Maria/Mria
Ysabel/Isabel
Query Expansion
Nadezhda/
Nadezda
Mikhailovna/
Mihajlovna
Baranskaya/
Baranskaja/
Baransky
106
Name Retrieval Technologies
Transposition
Nam, Bag Cho
Cho, Nam Bag
Bag, Cho Nam
Prefix Deletion
Nadir, Muhamad
Azia
Inverted Surnames
Vargas Gomez, Maria
Deleted Surname
Gomez, Maria
Nicknames
Nadya
Baransky
108
110
Nam, Park Cho
Narm, Bak Joe
Cho, Pak An
Cho, Nam Li
Gomez Vargas, Maria Ysabel
Bargas, Mario Ysidro
Gomez, Ma Isabel
Vargas Sosa, Maria Ysabel
Nadezhda Mikhailovnz Baranskaya/
Natasha Brownskaya/
Nadya Baransky/
Nadezka Mihajlovna Baranskaja/
Natalya brodsky
Abdul Nadir, Mho
Abdunnadir, Mohamed Aziz
Abd El Natheer, Mahmoud
Name Resource Library
100
112
Precision Filter and Sorting
Cont'd on FIG.1-1
Cont'd on FIG.1-1

FIG.1

Page 36 of 125

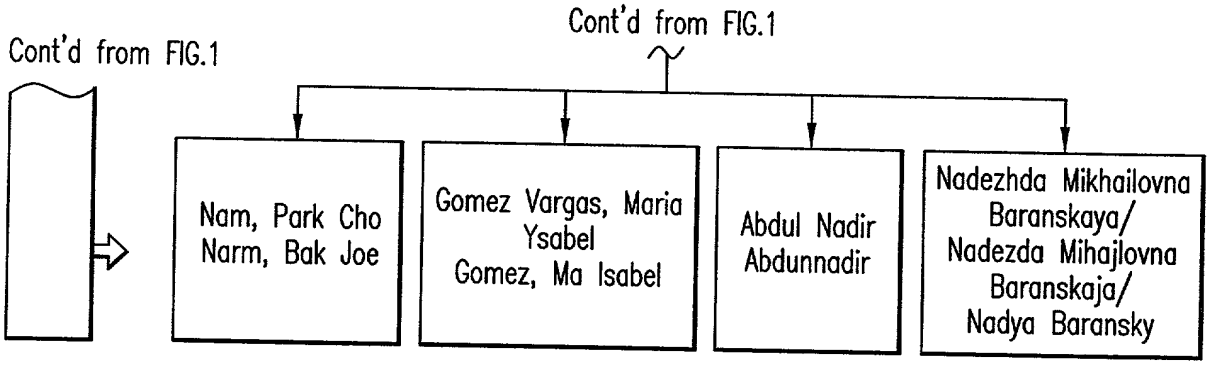

FIG. 1-1

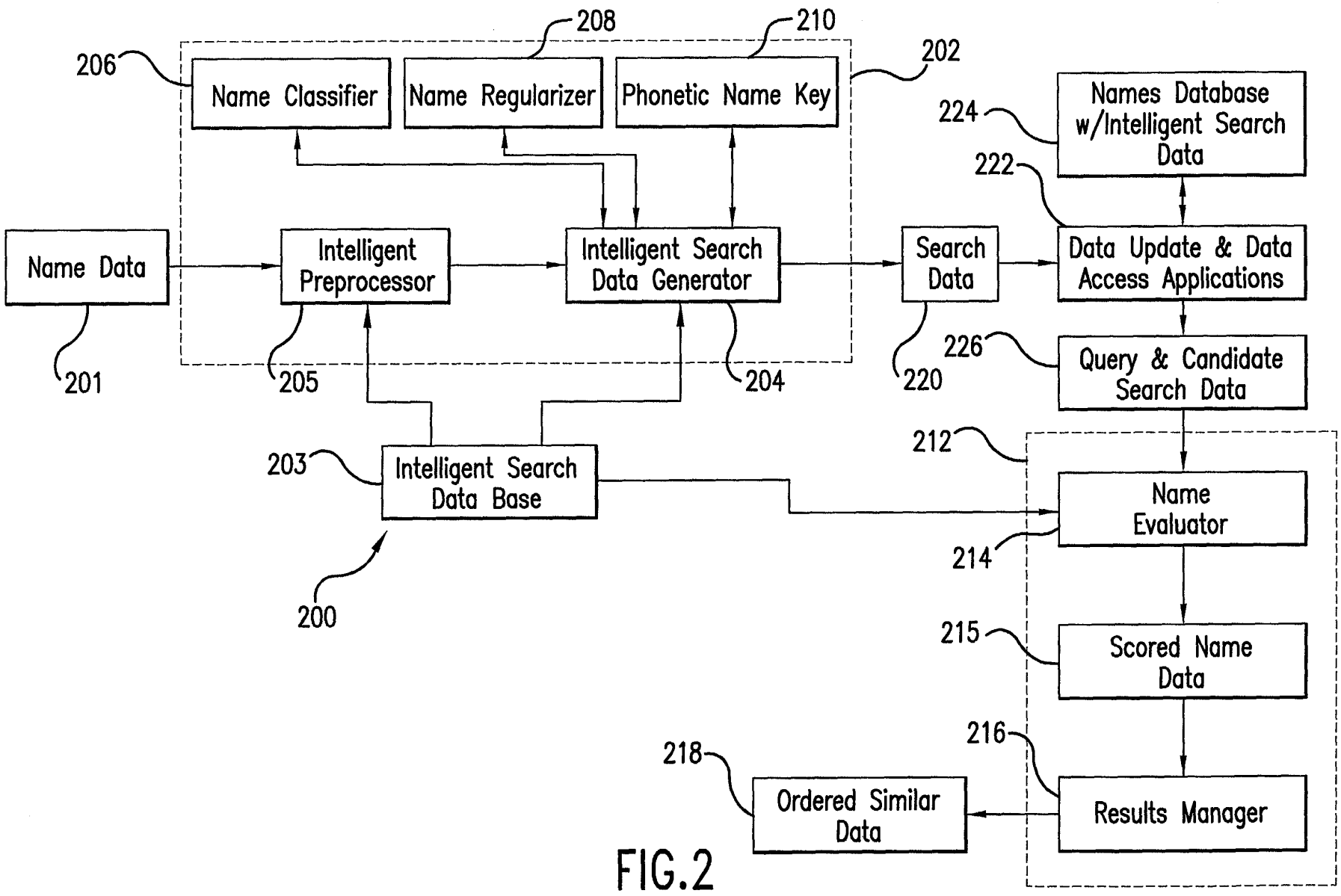
206
208
210
202
Name Classifier
Name Regularizer
Phonetic Name Key
224
Names Database w/Intelligent Search Data
222
Name Data
Intelligent Preprocessor
Intelligent Search Data Generator
Search Data
Data Update & Data Access Applications
201
205
204
220
226
Query & Candidate Search Data
212
203
Intelligent Search Data Base
Name Evaluator
214
200
215
Scored Name Data
216
218
Ordered Similar Data
Results Manager

FIG.2

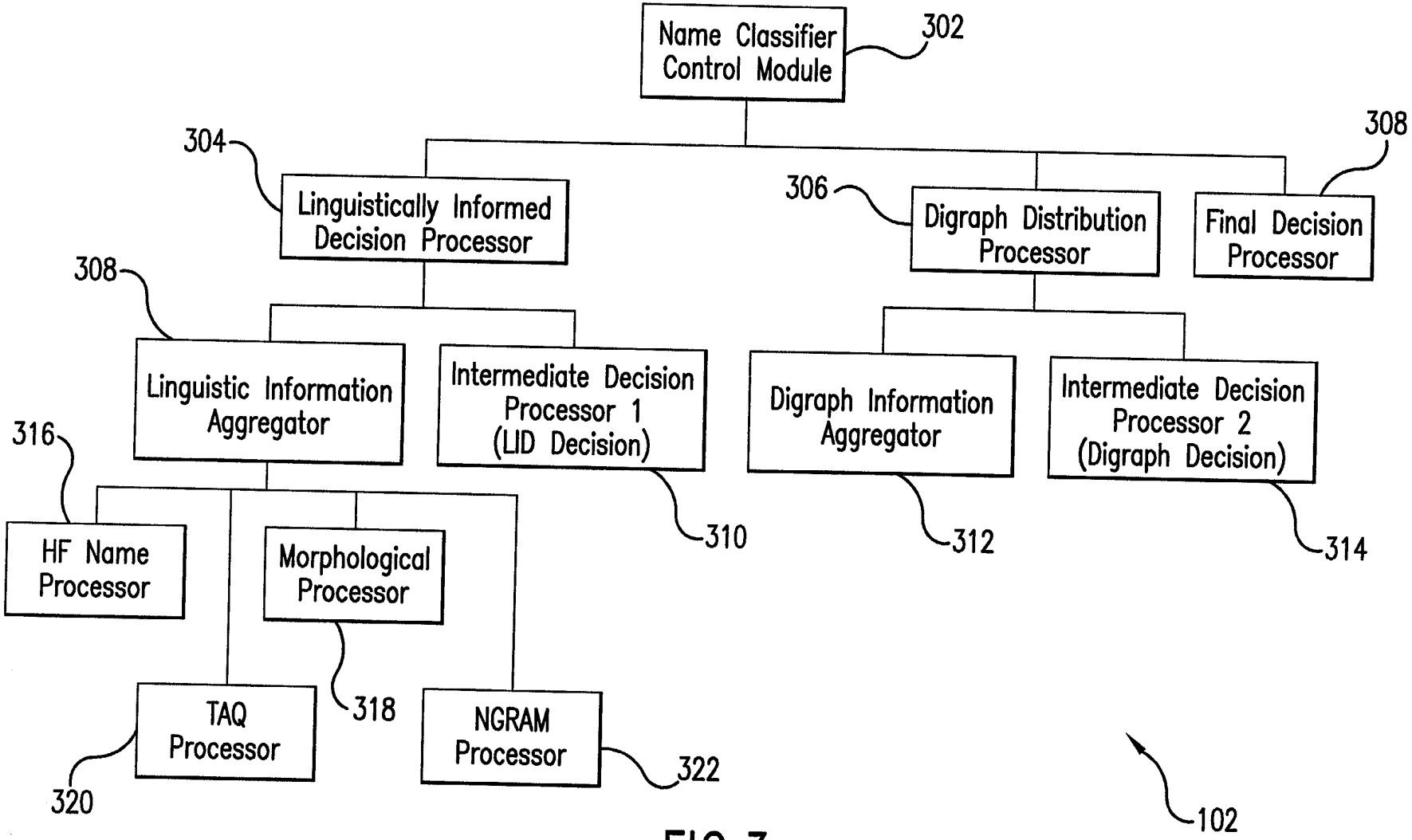
Name Classifier Control Module
302
304
Linguistically Informed Decision Processor
306
Digraph Distribution Processor
308
Final Decision Processor
308
Linguistic Information Aggregator
Intermediate Decision Processor 1 (LID Decision)
310
Digraph Information Aggregator
312
Intermediate Decision Processor 2 (Digraph Decision)
314
316
HF Name Processor
Morphological Processor
318
TAQ Processor
320
NGRAM Processor
322
102

FIG.3

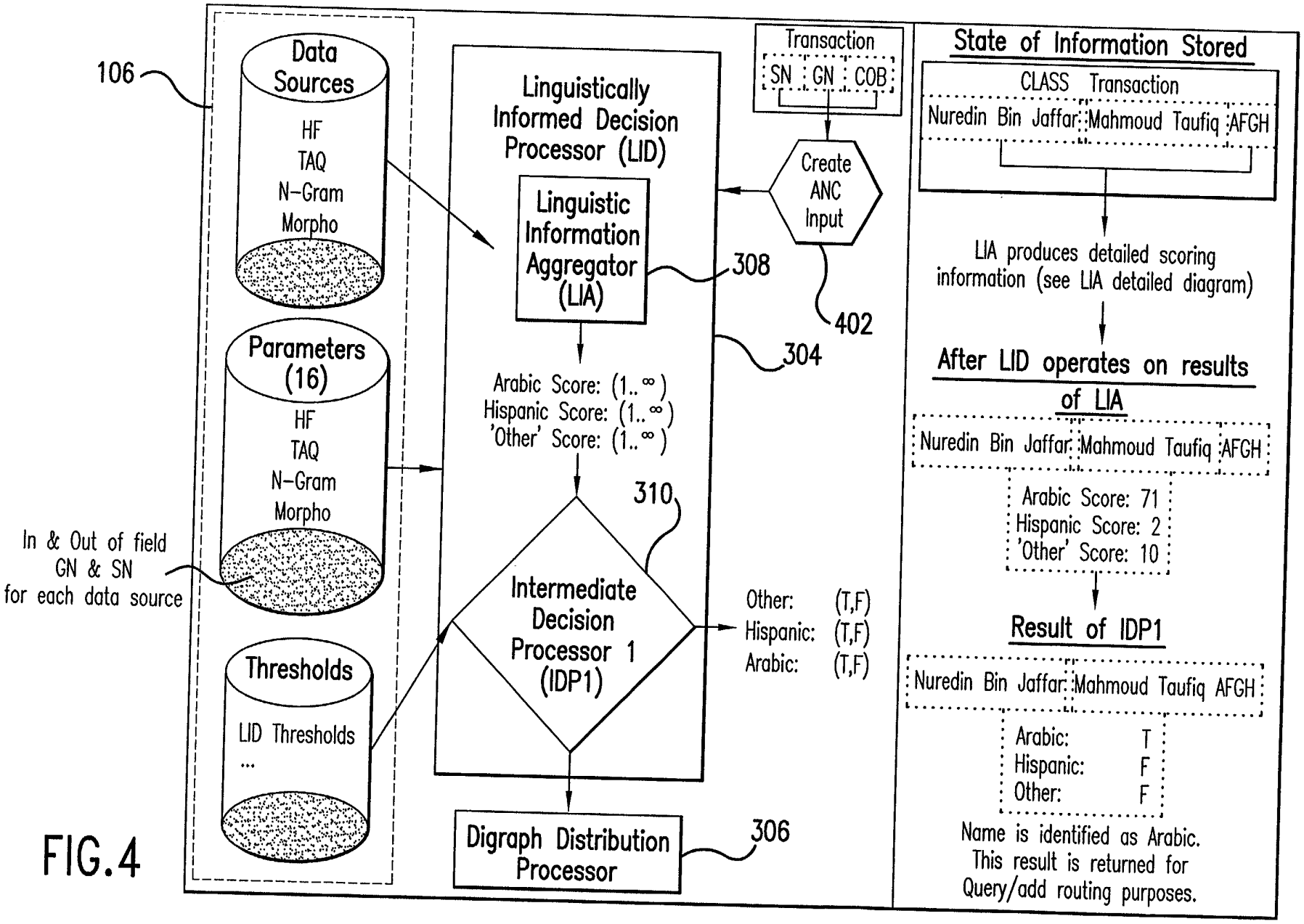
106
Data Sources
HF
TAQ
N-Gram
Morpho
Parameters (16)
HF
TAQ
N-Gram
Morpho
In & Out of field GN & SN for each data source
Thresholds
LID Thresholds
...
Linguistically Informed Decision Processor (LID)
Linguistic Information Aggregator (LIA)
308
Arabic Score: (1..∞)
Hispanic Score: (1..∞)
'Other' Score: (1..∞)
310
Intermediate Decision Processor 1 (IDP1)
Other: (T,F)
Hispanic: (T,F)
Arabic: (T,F)
304
Digraph Distribution Processor
306
Transaction
SN GN COB
Create ANC Input
402
State of Information Stored
CLASS Transaction
Nuredin Bin Jaffar Mahmoud Taufiq AFGH
LIA produces detailed scoring information (see LIA detailed diagram)
After LID operates on results of LIA
Nuredin Bin Jaffar Mahmoud Taufiq AFGH
Arabic Score: 71
Hispanic Score: 2
'Other' Score: 10
Result of IDP1
Nuredin Bin Jaffar Mahmoud Taufiq AFGH
Arabic: T
Hispanic: F
Other: F
Name is identified as Arabic.
This result is returned for Query/add routing purposes.

FIG.4

LIA Data Stores

High Frequency			
H	S	Garcia	3
H	S	Salazar	2
H	S	Sambrano	1
A	B	Mahmoud	4
A	B	Jaffar	2
O	S	Silvestri	1

N-Grams				
H	S	S	ndez	3
H	S	G	far	1
H	S	P	bous	1
A	B	S	fig	2
O	S	E	agio	1
O	S	I	ahmo	5

Morphology				
A	B	S	addin	3
A	B	S	edin	3
A	B	S	uddin	3
O	S	P	etto	1
O	S	S	etti	2
O	S	S	ini	1

TAQ			
H	S	de	1
H	S	la	1
H	S	las	1
A	B	bin	3
O	B	el	1
O	B	lo	1

Digraph Distribution Processor Data Stores

Digraphs		
A	ez	-1.0422
A	nt	22.8733
A	bd	38.7221
H	bd	1.0572
H	ez	42.5947
H	ri	16.1242

Trigraphs		
A	ez#	-10.0422
A	#nt	-48.1743
A	bd#	48.4551
H	bd#	-32.1742
H	ez#	47.5327
H	#ri	11.1242

Digraph Processor Parameters	
A	44.2331
H	-32.8765

LID Parameters			
HFNAME	S	10	8
HFNAME	G	8	6
TAO	S	5	3
TAO	G	4	2
MORPHOLOGY	S	5	3
MORPHOLOGY	G	4	2
NGRAM	S	4	3
NGRAM	G	3	2

Threshold Parameters				
A	50	1.4532	10	8
H	73	20.5000	5	6
O	38	(null)	100	3

FIG.5

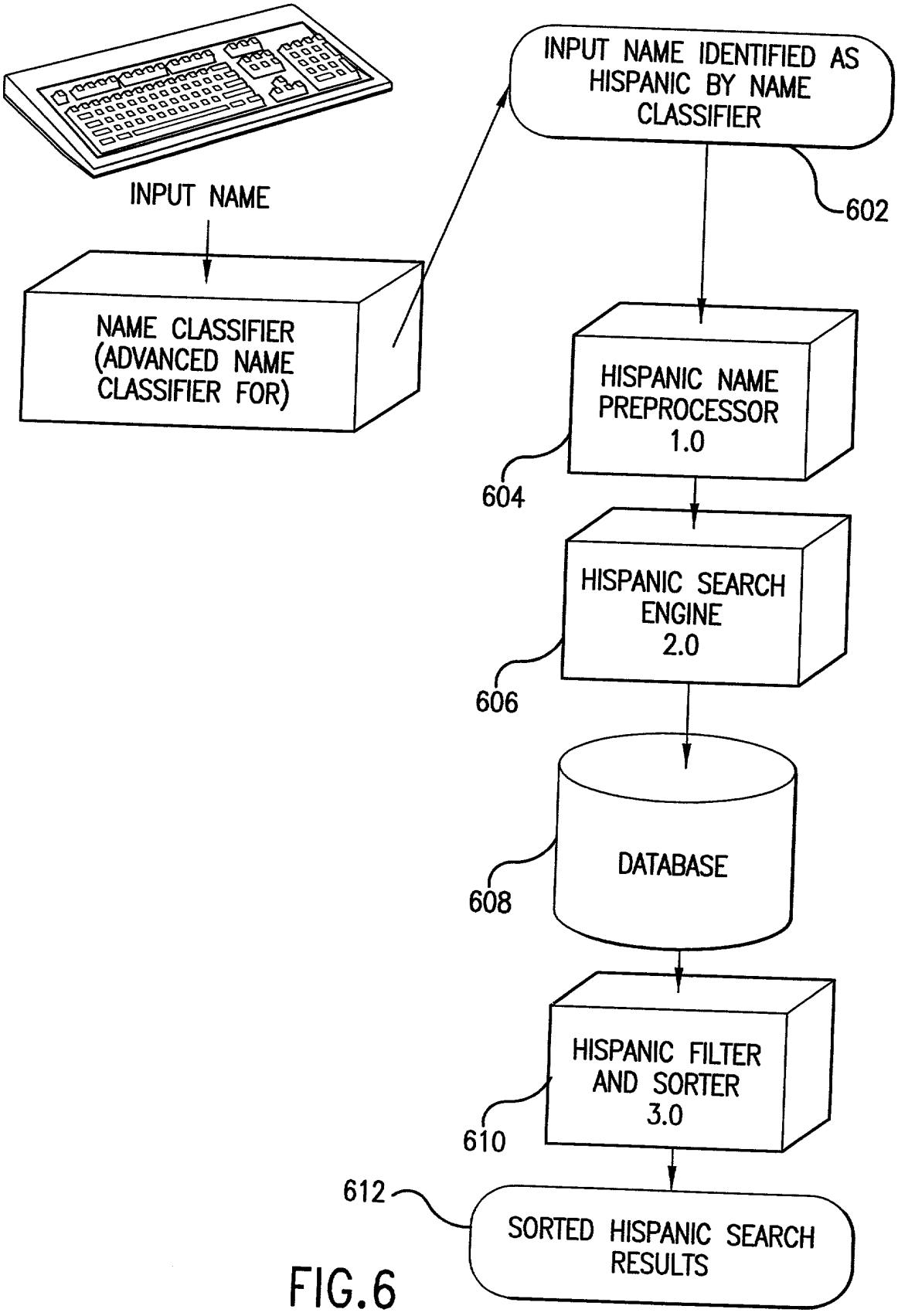
INPUT NAME
NAME CLASSIFIER (ADVANCED NAME CLASSIFIER FOR)
INPUT NAME IDENTIFIED AS HISPANIC BY NAME CLASSIFIER
602
HISPANIC NAME PREPROCESSOR 1.0
604
HISPANIC SEARCH ENGINE 2.0
606
DATABASE
608
HISPANIC FILTER AND SORTER 3.0
610
612
SORTED HISPANIC SEARCH RESULTS

FIG.6

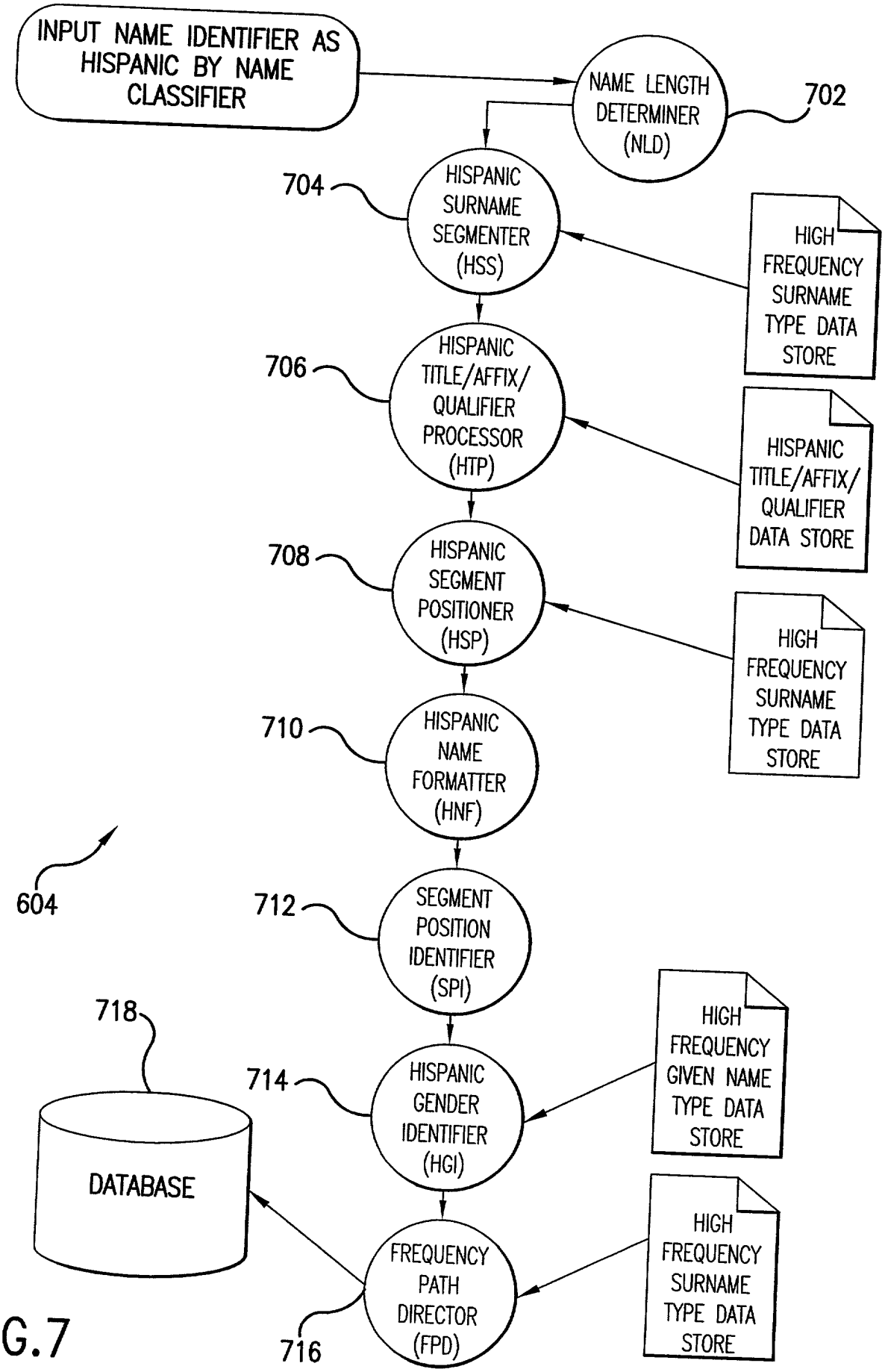
INPUT NAME IDENTIFIER AS HISPANIC BY NAME CLASSIFIER
NAME LENGTH DETERMINER (NLD)
702
704
HISPANIC SURNAME SEGMENTER (HSS)
HIGH FREQUENCY SURNAME TYPE DATA STORE
706
HISPANIC TITLE/AFFIX/ QUALIFIER PROCESSOR (HTP)
HISPANIC TITLE/AFFIX/ QUALIFIER DATA STORE
708
HISPANIC SEGMENT POSITIONER (HSP)
HIGH FREQUENCY SURNAME TYPE DATA STORE
710
HISPANIC NAME FORMATTER (HNF)
604
712
SEGMENT POSITION IDENTIFIER (SPI)
718
714
HISPANIC GENDER IDENTIFIER (HGI)
HIGH FREQUENCY GIVEN NAME TYPE DATA STORE
DATABASE
FREQUENCY PATH DIRECTOR (FPD)
716
HIGH FREQUENCY SURNAME TYPE DATA STORE

FIG.7

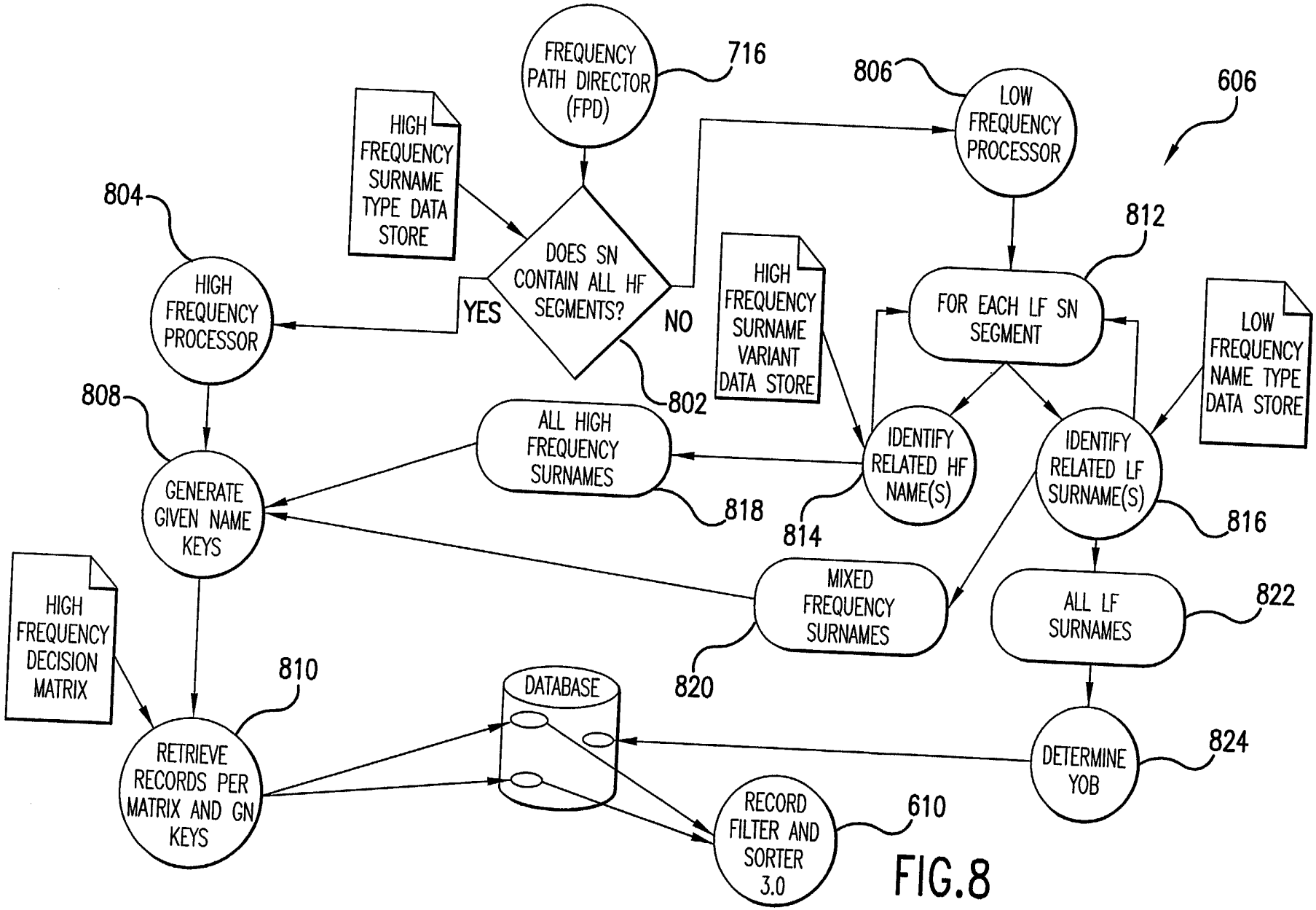
FREQUENCY PATH DIRECTOR (FPD)
716
HIGH FREQUENCY SURNAME TYPE DATA STORE
DOES SN CONTAIN ALL HF SEGMENTS?
YES
NO
802
804
HIGH FREQUENCY PROCESSOR
806
LOW FREQUENCY PROCESSOR
606
812
FOR EACH LF SN SEGMENT
HIGH FREQUENCY SURNAME VARIANT DATA STORE
LOW FREQUENCY NAME TYPE DATA STORE
808
ALL HIGH FREQUENCY SURNAMES
818
IDENTIFY RELATED HF NAME(S)
814
IDENTIFY RELATED LF SURNAME(S)
816
GENERATE GIVEN NAME KEYS
MIXED FREQUENCY SURNAMES
820
ALL LF SURNAMES
822
HIGH FREQUENCY DECISION MATRIX
810
DATABASE
RETRIEVE RECORDS PER MATRIX AND GN KEYS
DETERMINE YOB
824
RECORD FILTER AND SORTER 3.0
610

FIG.8